Statistical analysis of type S thermocouple measurements on the International Temperature Scale of 1990

W.F. Guthrie, M.C. Croarkin, G.W. Burns, and G.F. Strouse
National Institute of Standards and Technology (NIST)
Gaithersburg, MD 20899, USA

P. Marcarino and M. Battuello
Istituto di Metrologia 'G. Colonnetti' (IMGC)
Torino, Italy

H.K. Lee, J.C. Kim, K.S. Gam, and C. Rhee
Korea Research Institute of Standards and Science (KRISS)
Taejon, Republic of Korea

M. Chattle
National Physical Laboratory (NPL)
Teddington, UK

M. Arai and H. Sakurai
National Research Laboratory of Metrology (NRLM)
Ibaraki, Japan

A.I. Pokhodun, N.P. Moiseeva, and S.A. Perevalova
D.I. Mendeleyev Institute for Metrology (VNIIM)
St. Petersburg, Russia

M.J. de Groot
Van Swinden Laboratory (VSL)
Delft, The Netherlands

Jipei Zhang, Kai Fan, and Shuyuan Wu
Shanghai Institute of Process Automation Instrumentation (SIPAI)
Shanghai, PRC

ABSTRACT

Data collected by scientists from eight national laboratories were used to estimate the differences between temperatures on the International Temperature Scale of 1990 and the International Practical Temperature Scale of 1968, Amended Edition of 1975 between 630.615 °C and 1064.18 °C. A new reference function for type S thermocouples was determined for the temperature range −50 °C to 1064.18 °C. The new reference function was combined with two other reference functions for temperatures between 1064.18 °C and 1768.1 °C to provide a complete set of reference functions on ITS-90 for type S thermocouples. This paper describes the modeling procedures used to determine the reference functions and statistical analyses used to estimate differences between the two temperature scales. Issues addressed include: variability within laboratories, form of the new reference functions, and the uncertainty associated with the reference functions.

SUBJECT INDEX: International Temperature Scale of 1990 (ITS-90), Noble metal thermocouple thermometers, Statistical analysis.

INTRODUCTION

This paper describes how type S thermocouple measurement data were used to estimate the difference between temperatures on the International Temperature Scale of 1990 (ITS-90) (1) and the International Practical Temperature Scale of 1968, Amended Edition of 1975 (IPTS-68) (2) for temperatures in the range 630.615 °C to 1064.18 °C. The development of a new reference function for type S thermocouples for the temperature range −50 °C to 1064.18 °C is also described. The data were collected by eight national laboratories as part of an international experiment to develop ITS-90 based thermocouple reference functions (3).

Throughout the paper, t_{90} and t_{68} denote temperatures measured in degrees Celsius on the ITS-90 and IPTS-68, respectively. The difference between the two scales is denoted by $\Delta t = t_{90} - t_{68}$.

ESTIMATION OF $\Delta t = t_{90} - t_{68}$

The data used for estimating the temperature differences between the ITS-90 and the IPTS-68 are shown in Figure 1. Data from each lab are denoted by a unique symbol. The temperature differences were obtained by taking the differences of the temperatures measured directly on the ITS-90 and the corresponding IPTS-68 temperatures calculated using the associated emf measurement and the IPTS-68 defining quadratic function for each thermocouple. The details of the temperature difference computations are discussed in (3).

Figure 1 makes it clear that there are systematic errors between sets of temperature differences derived from measurements made by different laboratories using different thermocouples. In addition to the systematic errors in each set of temperature differences there are, of course, random measurement errors as well.

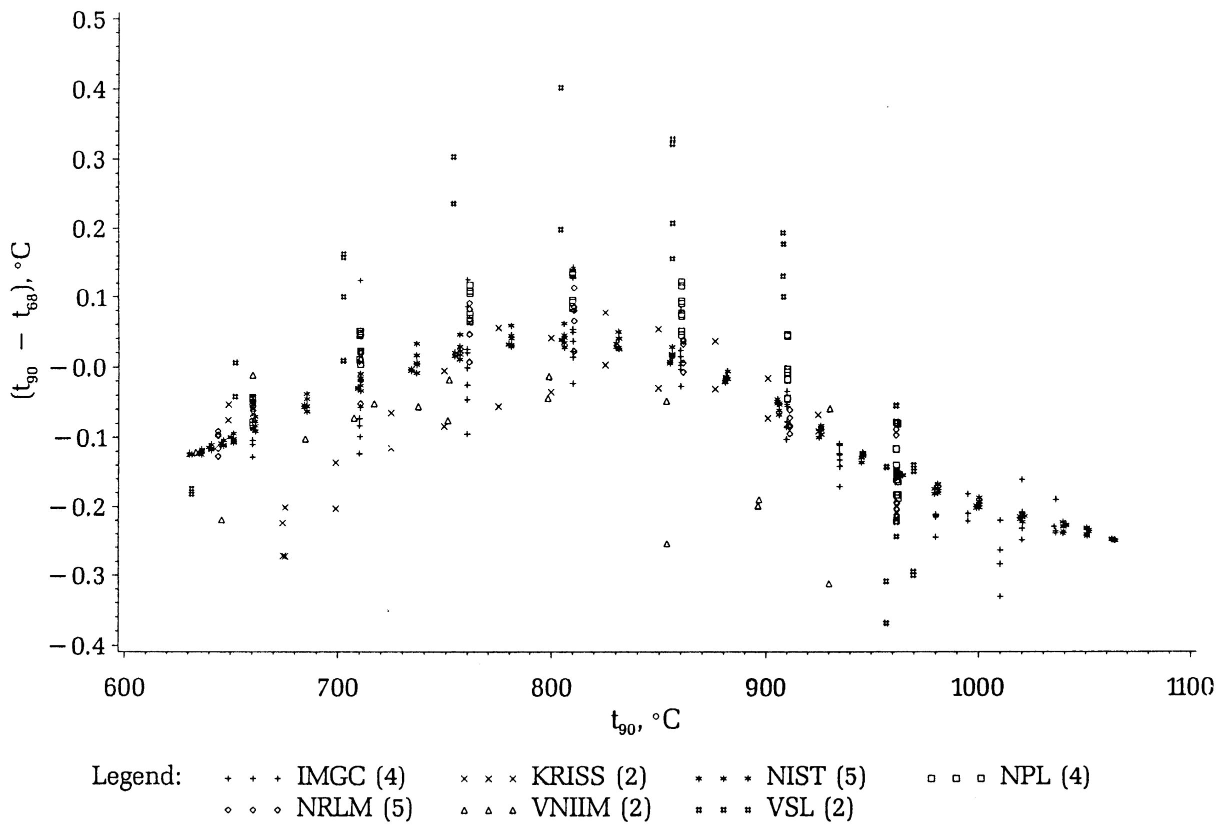

Figure 1: Measured temperature differences vs. t_{90}. The number of thermocouples tested by each laboratory is shown in parentheses.

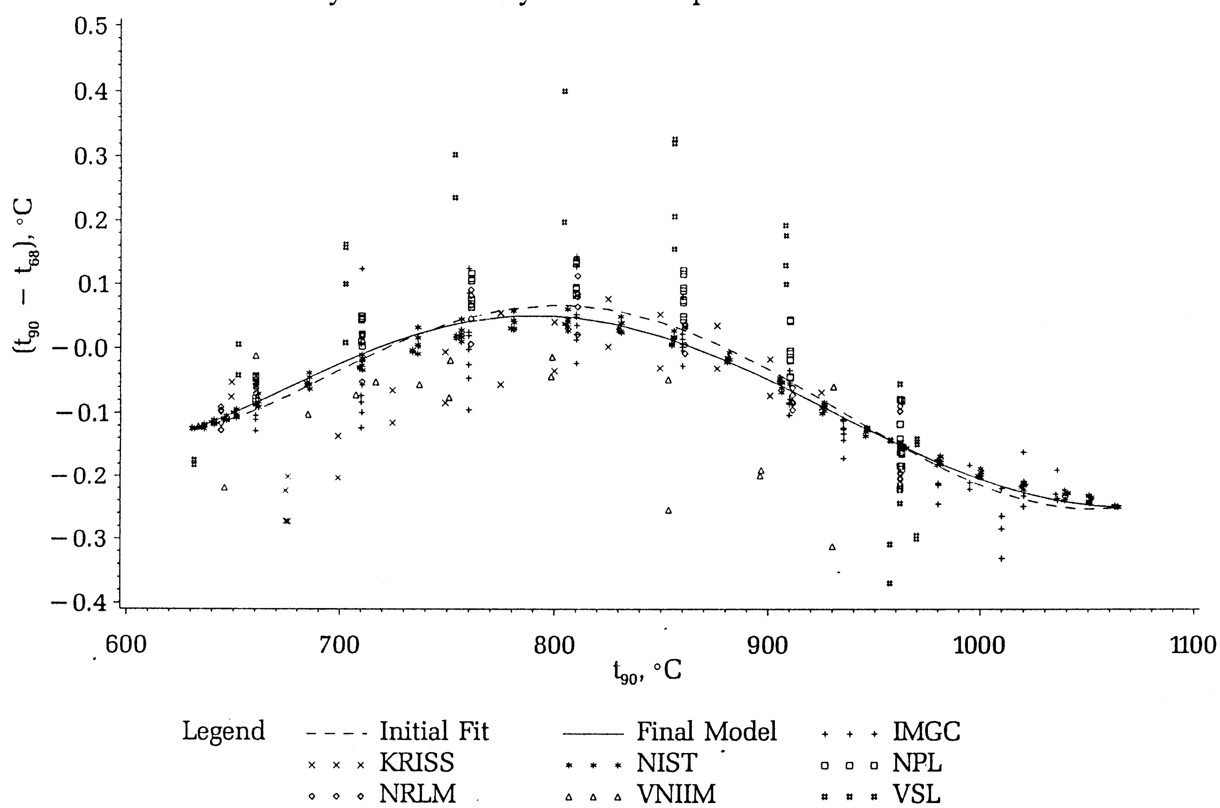

Figure 2: Measured temperature differences, initial model; and final model vs. t_{90}.

Because the data have a complicated error structure, and because there is no way to determine which set of temperature differences is closest to the actual difference between the two temperature scales, the difference between the scales is estimated by the consensus of data from different sources. In this paper, we assume that the systematic errors associated with each lab/thermocouple combination are randomly distributed across laboratories and thermocouples, and that the distribution of the systematic errors is centered at zero. Under these assumptions, the consensus value gives an unbiased estimate of the difference between temperatures on the ITS-90 and the IPTS-68.

To develop a consensus model for the difference between temperatures on the two scales, we fitted a 5^{th} degree polynomial in t_{90} to the temperature difference data using least squares regression. The polynomial model was chosen because no theoretical model relating Δt to t_{90} is known. The 5^{th} degree polynomial provided a reasonable summary of the data. Polynomials of lower degree did not adequately describe the difference in temperatures on the two scales as a function of t_{90}. Polynomials of higher degree did not significantly improve the fit to the data, so use of additional terms in the model was unwarranted. The regression function is shown in Figure 2 and is labeled 'Initial Model' in the legend.

To reduce the influence of isolated data points on the estimation procedure, we standardized the residuals and then weighted each point in the data set using weights inversely related to the magnitude of the standardized residual. Then, we refitted the model using weighted least squares regression to obtain estimates of the regression parameters that were less prone to effects of outlying data. This process of weighting the residuals after fitting the model and then refitting the model was repeated until the regression function converged to a final model, which better reflects the consensus of the data than the initial model does. This repeated fitting/weighting/fitting procedure is known as iteratively reweighted least squares regression (IRLS regression) in statistical literature (4).

Equation 1 shows how the i^{th} standardized residual is computed for the j^{th} iterative fit.

$$u_{i,j} = \frac{d_i - \hat{d}_{i,j}}{cS_j} \tag{1}$$

The standardized residuals are denoted by $u_{i,j}$, the measured temperature difference by d_i, and the predicted temperature difference by $\hat{d}_{i,j}$. The denominator of $u_{i,j}$ is the product of a tuning constant, c, and S_j, a robust measure of the variability of the residuals. The median of the absolute residuals was used as the measure of residual variability, S_j, for this analysis. The tuning constant chosen was $c = 6$. For this experiment $i = 1, 2, \ldots, 466$ and $j = 1, 2, \ldots, 5$.

We used Tukey's biweight function (4) to compute the weights for each data point.

$$w(u_{i,j}) = \begin{cases} (1 - u_{i,j}^2)^2 & \text{if } |u_{i,j}| \le 1 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

The biweight function produces weights between 0 and 1. Observations near the regression function receive a weight close to 1 and the weights progressively decrease as the observations get further away from the regression function. Points that are more than c units away from the regression function (measured in scale units of size S_j) receive a weight of 0 and are omitted from the next model fit. However, these points could receive positive weight again in a later step.

Figures 3 and 4 show some of the details of the IRLS regression analysis. Figure 3 shows the weights that were used for the final fit. In the final fit, 414 out of 466 data points had weights greater than zero.

Figure 4 shows the convergence of the model, as measured by the sample residual standard deviation, to its final form. The sample residual standard deviation from each fit is plotted versus iteration number in this figure. The sample residual standard deviation from the initial fit is inflated by outlying points in the data set. As the data are reweighted in successive steps, the residual standard deviation shrinks towards a value which more accurately reflects the random measurement error in the data. When all of the data have been appropriately weighted, the sample residual standard deviation remains unaltered by further reweighting and refitting. In this experiment, the initial model had a residual standard deviation of 62.6 m°C. After five iterations, the residual standard deviation converged to 17.5 m°C.

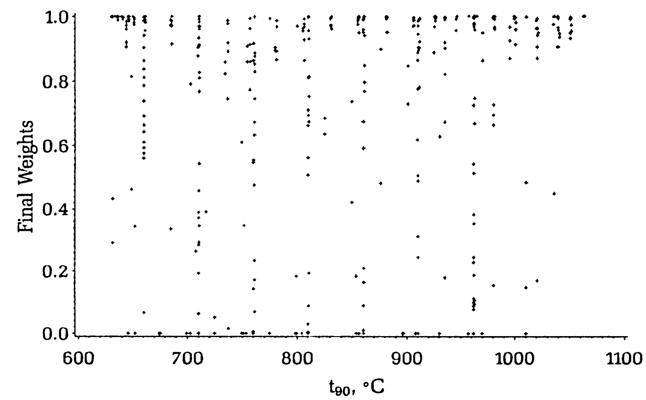

Figure 3: Weights used to fit the final model.

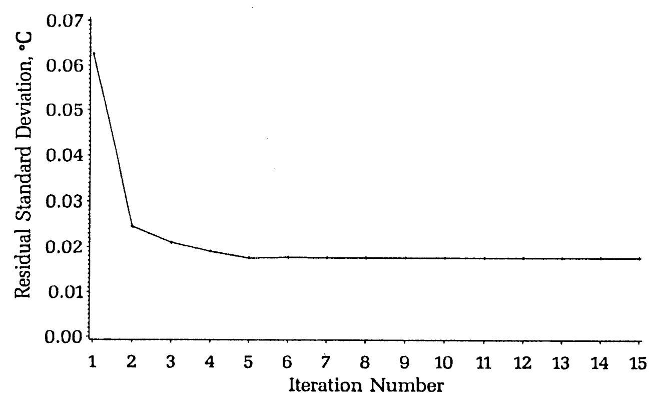

Figure 4: Convergence of iterative models to the final model as measured with the residual standard deviation.

As mentioned above, the final regression function should describe the consensus of data and should not be affected by outlying data points in the data set. The final model from the iteratively reweighted least squares analysis is also shown in Figure 2. Comparing the initial and final models with the data from the the different laboratories makes it clear that the iterative weighting and fitting reduced the influence of the data isolated from the bulk of the measured temperature differences.

In addition to using regression analysis techniques that reduce the effect of outliers in the data, the function $\Delta t(t_{90})$ was constrained to match the adopted temperature differences (5) of −0.125 °C, −0.15 °C, and −0.25 °C at t_{90} = 630.615 °C, t_{90} = 961.78 °C, and t_{90} = 1064.18 °C, respectively. $\Delta t(t_{90})$ matches the adopted temperature differences to at least 0.1 m°C at these three points.

$\Delta t(t_{90})$ was not constrained to have the same first derivative (or slope) that the published scale difference (1,5) has at either t_{90} = 630.615 °C or at t_{90} = 1064.18 °C. The discontinuity in the first derivative at t_{90} = 630.615 °C is approximately 0.14%. This discontinuity is smaller than the 0.51% reported in (6), but this estimate of the discontinuity has a magnitude that is similar to estimates from earlier experiments (7,8,9). The discontinuity at t_{90} = 1064.18 °C is about 0.008%, which is negligible relative to the errors in the determination of $\Delta t(t_{90})$.

The final regression model which describes the temperature differences between the ITS-90 and the IPTS-68 is

$$\begin{aligned}\Delta t(t_{90}) = \; & (7.8687209 \times 10^{1}) \\ & -(4.7135991 \times 10^{-1})t_{90} \\ & +(1.0954715 \times 10^{-3})t_{90}^{2} \\ & -(1.2357884 \times 10^{-6})t_{90}^{3} \\ & +(6.7736583 \times 10^{-10})t_{90}^{4} \\ & -(1.4458081 \times 10^{-13})t_{90}^{5}.\end{aligned} \tag{3}$$

Figure 5 compares the differences between temperatures on the ITS-90 and the IPTS-68 as published in (1) and as estimated from this experiment. In the table of differences in (1), the maximum difference between the two temperature scales in the range t_{90} = 630.615 °C to t_{90} = 1064.18 °C is 0.36 °C. The maximum is attained at t_{90} = 760 °C, t_{90} = 770 °C, t_{90} = 780 °C (this is a non-unique maximum). The estimate of the difference between the two scales obtained from this experiment has a maximum of 0.051 °C at t_{90} = 790.916 °C. The maximum disagreement between these two estimates of Δt is 0.318 °C at t_{90} = 760 °C.

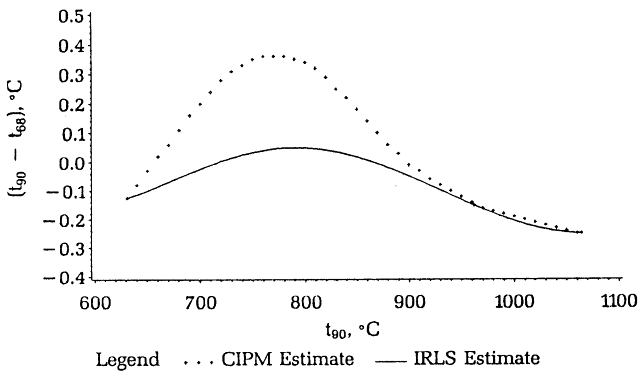

Figure 5: Comparison of IRLS estimate of ($t_{90} - t_{68}$) with previously published CIPM estimate of ($t_{90} - t_{68}$).

Clearly there is significant disagreement between the table of differences in (1) and the data and estimate of Δt presented here. This disagreement should be taken into consideration in any use of the temperature differences between the ITS-90 and the IPTS-68. The estimate of Δt presented here is based on results from a large experiment with the participation of many laboratories. The polynomial describing the temperature differences between the two scales was derived using robust regression techniques which protect against biases caused by isolated, outlying data points. Assuming that the systematic errors present in the data are randomly distributed with zero mean across different lab/thermocouple combinations, the estimate of Δt presented here should be an accurate reflection of the actual difference between temperatures on the ITS-90 and the IPTS-68.

REFERENCE FUNCTION ON ITS-90 FOR TYPE S THERMOCOUPLES

The reference function for type S thermocouples based on the IPTS-68 consisted of four functions, $g_i(t_{68})$, $(i = 1, \ldots, 4)$, with breaks at t_{68} = 630.74 °C, t_{68} = 1064.43 °C and t_{68} = 1665 °C (7). The reference function based on the ITS-90 consists of a single function, $f_1(t_{90})$, in the range −50 °C to 1064.18 °C and two functions, $f_2(t_{90})$ and $f_3(t_{90})$, in the range above 1064.18 °C (3). The latter are the result of substitution of $t_{90} - \Delta t$ for t_{68} in the IPTS-68 reference functions with modifications as described in Burns, Strouse, et al. (3).

The purpose of this section is to describe the derivation of $f_1(t_{90})$ from the data of the current experiment. The experiment was undertaken because satisfactory reference functions on ITS-90 in the range −50 °C to 1064.18 °C could not be produced by substitution in $g_1(t_{68})$ and $g_2(t_{68})$. Furthermore, an assessment of uncertainty was not possible without a new experiment because the earlier data on type S thermocouples were not archived. Analysis of the present data produced a single function based on the ITS-90 to replace $g_1(t_{68})$ and $g_2(t_{68})$.

The only candidates for determining $f_1(t_{90})$ were NIST thermocouples, S1, S2, S3, S4, and S5 because they were the only thermocouples measured over the full range between −50 °C and 1070 °C. Thermocouple S3 was not considered because it was annealed in a different manner from the other thermocouples, and it was not as stable. There were no repeated runs with thermocouples S1 and S2; consequently, there were fewer data for them than for S4 and S5. Thermocouple S5 was preferred over S4, although S4 was made from the same material that was the basis for the IPTS-68 reference function, because the S5 emf-temperature relationship agreed more closely with the IPTS-68 reference function. Also, S5 was more thermoelectrically homogeneous; its fixed point data agreed more closely with its comparison data than was the case for S4. The data for all five NIST thermocouples and thermocouples from other laboratories are compared in (3).

The data on S5 consisted of a relatively few fixed point determinations and a large number of comparisons with standard platinum resistance thermometers calibrated in accordance with the ITS-90. Comparison measurements were made in five apparatuses as shown in the table below. Below 0 °C, comparisons were made at 10 °C intervals; above 0 °C, comparisons were made at roughly 20 °C intervals except in the neighborhoods of 630 °C and 1064 °C where temperatures were closely spaced. Measurements were made with both increasing and decreasing temperatures with multiple readings at each temperature; however, only data with increasing temperature contributed to the reference function. Data were reduced to four points at each nominal temperature, and the entire sequence was repeated on one or two occasions. The experimental plan for S5 is shown in Table I.

Table I: Experimental Plan for Thermocouple S5.

Apparatus	Range			Occasions
Cryostat	−50 °C	to	−10 °C	2
Water bath	0 °C	to	95 °C	2
Oil bath	95 °C	to	180 °C	1
Salt bath	275 °C	to	550 °C	1
Comparator	500 °C	to	1070 °C	2

The reference function in the range −50 °C to 1064.18 °C is an $8th$ degree polynomial, $f_1(t_{90})$, of the form

$$f_1(t_{90}) = \sum_{i=1}^{8} \beta_i t_{90}^{i} \tag{4}$$

where the coefficients, β_i $(i = 1, \ldots, 8)$, were initially determined from a least squares fit, $p(t_{90})$, to 444 measurements of emf as a function of temperature on the ITS-90. The residual standard deviation was 0.0629 μV with 436 degrees of freedom. The residuals from the fit are shown in Figure 6.

An $8th$ degree polynomial was chosen as the reference function for the following reasons. Lower degree polynomials did not prove adequate in that the residuals from the fit, which should be randomly distributed about zero, showed cyclic structure. A test statistic for each β coefficient was formed as the ratio of β to its standard error. All coefficients for the $8th$ degree polynomial tested significantly different from zero at the 0.0001 significance level, indicating the need for at least an $8th$ degree fit. Furthermore, the $8th$ degree fit proved successful in removing cyclic structure across different pieces of equipment, although retaining some structure within an apparatus. Polynomials of higher degree were successful to some extent in removing within apparatus structure from the residuals; however, the reduction in the mean squared error, achieved by going from an $8th$

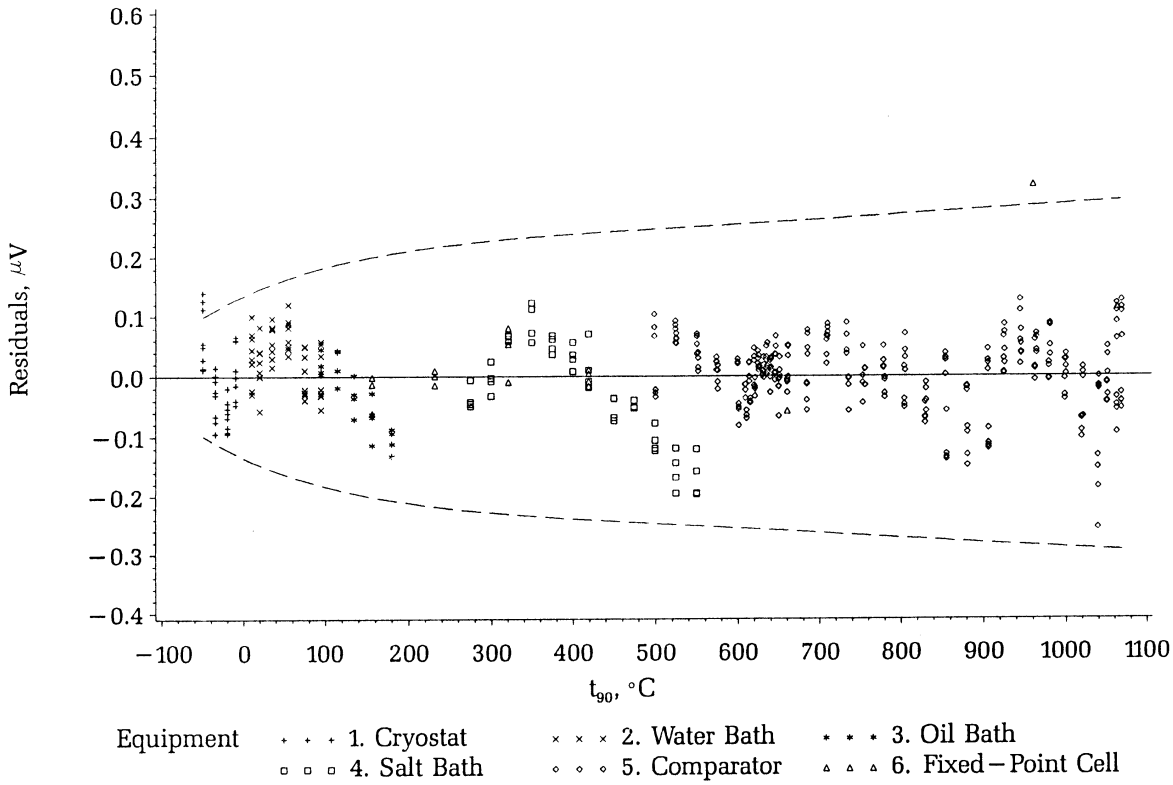

Figure 6: Residuals from an 8th degree polynomial model vs. t_{90} for NIST thermocouple S5. The dashed lines are ±25 m°C bounds on the residuals in terms of emf.

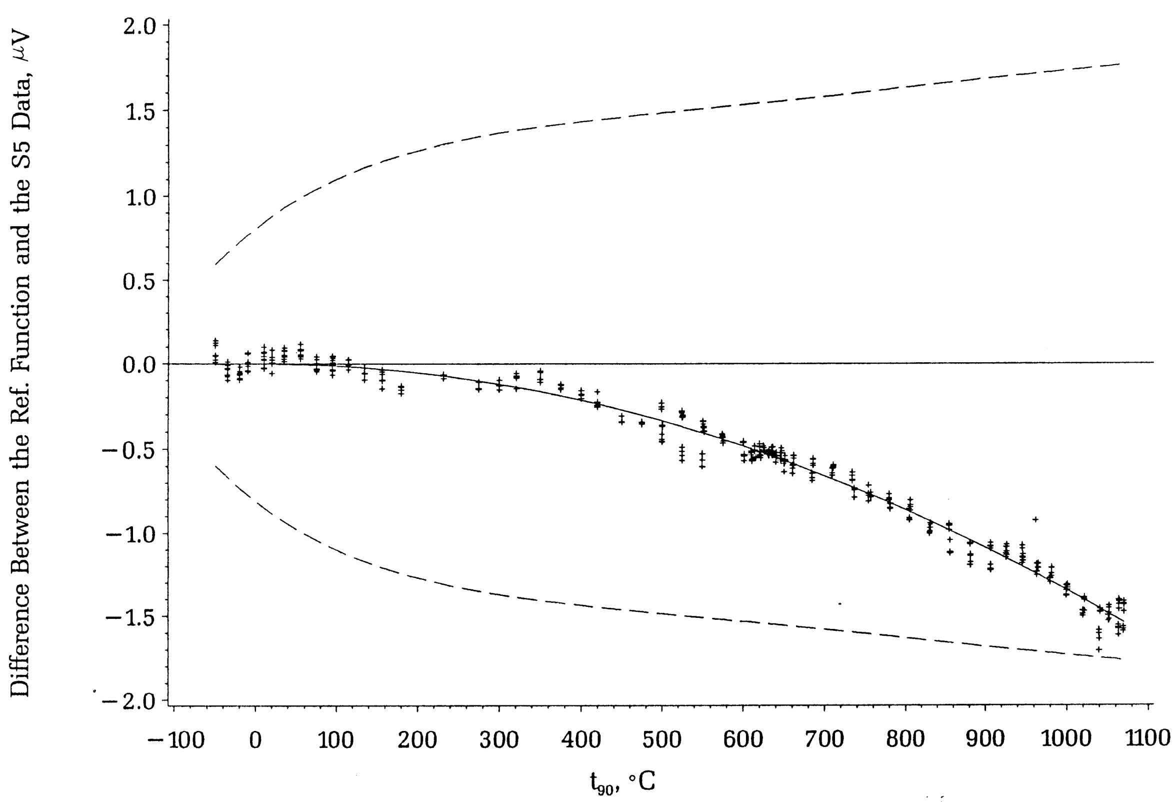

Figure 7: Deviation of S5 data and the polynomial fit to the data from the reference function as published in (3). The solid line indicates the 8th degree polynomial fit. The dashed lines indicate deviations equivalent to ±0.15 °C.

degree polynomial to a higher degree polynomial, did not prove substantial enough to warrant increasing the order of the reference function. Also, the structure in the residuals suggests that small changes occurred in the thermocouple with increasing temperature. These changes should not be reflected in the reference function. Therefore, the model does not account for this type of structure.

The random component of uncertainty for $p(t_{90})$ is calculated using Working-Hotelling confidence bands (10). The upper and lower 95% confidence bands at temperature t_h are $p(t_h) \pm v(t_h)$ where

$$v(t_h) = \sqrt{8F_{0.95}(8,436)}s_h. \tag{5}$$

The critical value $F_{0.95}(8,435) = 1.96$ is the upper 95 percent point of the F distribution with 8 and 436 degrees of freedom, and s_h is the standard deviation of $p(t_h)$ at temperature t_h. The Working-Hotelling confidence bands are appropriate for unlimited use of the reference function. Representative values are shown in Table II.

Table II: **Random uncertainties (μV) for $p(t_{90})$ from 95% Working-Hotelling confidence bands**

t °C	$p(t_{90})$	$v(t_{90})$
-50.00	-235.56	0.07
0.00	0.00	0.00
100.00	645.90	0.04
200.00	1440.73	0.04
300.00	2322.92	0.05
400.00	3259.14	0.04
500.00	4232.96	0.03
600.00	5238.21	0.02
700.00	6274.59	0.03
800.00	7344.12	0.03
900.00	8448.15	0.03
1000.00	9585.75	0.04
1064.18	10332.68	0.05

The final reference function, $f_1(t_{90})$, is derived from an adjustment of the quadratic coefficient, β_2', of $p(t_{90})$ to make $f_1(t_{90}) = f_2(t_{90})$ at $t_{90} = 1064.18$ °C. The adjusted coefficient is

$$\beta_2 = \frac{f_2(1064.18) - (p(1064.18) - \beta_2' 1064.18^2)}{(1064.18)^2}. \tag{6}$$

The magnitude of the adjustment at 1064.18 °C is 1.52 μV. Figure 7 shows the deviation of the S5 data from the reference function $f_1(t_{90})$. The adjustment guarantees that the reference function is continuous at 1064.18 °C. The coefficients for $f_1(t_{90})$ are reported by Burns, Strouse et al. (3).

REFERENCES

1. Preston-Thomas H. (1990) The International Temperature Scale of 1990 (ITS-90), *Metrologia*, **27**, 3-10, 107 (erratum).

2. The International Practical Temperature Scale of 1968, Amended Edition of 1975 (1976) *Metrologia*, **12**, 7-17.

3. Burns, G. W., Strouse, G. F. et al. (1992) New Reference Function for Platinum-10% Rhodium Versus Platinum (Type S) Thermocouples Based on the ITS-90, This Proceedings (American Institute of Physics).

4. Mosteller, F. and Tukey, J. W. (1977) Data Analysis and Regression, Addison-Wesley Publishing Co., Inc.: Menlo Park, CA, 357-365.

5. Rusby, R. L. et al. (1991) Thermodynamic basis of the ITS-90, *Metrologia*, **28**, 9-18.

6. Bedford, R. W. et al. (1987) Direct measurement of the discontinuity in dt/dt_{68} at 630.74 °C in the IPTS-68, *Metrologia*, **23**, 197-205.

7. Bedford, R. E. et al. (1972) *Temperature, Its Measurement and Control in Science and Industry, Vol 4*, H. H. Plumb (ed.) (Instrument Society of America, Pittsburgh), 1585-1602.

8. Evans, J. P. and Wood, S. D. (1971) An Intercomparison of High Temperature Platinum Resistance Thermometers and Standard Thermocouples, *Metrologia*, **7**, 108-130.

9. Crovini, L. and Actis, A. (1978) Noise Thermometry in the Range 630-962 °C, *Metrologia*, **14**, 69-78.

10. Neter, J., Wasserman, W., and Kutner, M. (1983) Applied Linear Regression Models, Richard D. Irwin, Inc.: Homewood, IL, 244.